

**Amendments to the Specification:**

Please replace paragraph [0011] with the following rewritten paragraph:

[0011] The content of each of the content portions, documents or web pages making up the document collection are determined in step S30. The words on each content portions, document or web page are added to a word / document frequency matrix. The weights of the words are determined and a weighted word document frequency matrix is created. The weighting may use term frequency/inverse document frequency ( $TF \cdot IDF$ ), log of ~~log of~~ the term frequency,  $1 + (\log_{10} \text{ of the term frequency})$  or any other known or later developed technique of weighting. Control then continues to step S40.